RANDOM WALKS WITH k-WISE INDEPENDENT INCREMENTS

ITAI BENJAMINI, GADY KOZMA, AND DAN ROMIK

ABSTRACT. We construct examples of a random walk with pairwise-independent steps which is almost-surely bounded, and for any m and k a random walk with k-wise independent steps which has no stationary distribution modulo m.

1. INTRODUCTION

Consider a simulation of a simple random walk on a graph. How will the simulation be affected if the source of randomness is not truly random but only pseudo random in the following specific sense, the random bits are k-wise independent for some $k > 1$ and not independent as a family? The first question to ask is, does the pseudo walk converge to the same stationary measure? The most simple graph to consider might be a cycle of size m. This suggests the following problem: given a random walk $S_n = \sum_{i=1}^n X_i$, where the X_i are random signs, plus or minus 1 with equal probability, and the X_i's are k-wise independent, what can be said about the behavior of the partial sums and in particular modulo some fixed number m? It turns out that there is a fundamental difference between the cases $k = 2$ and $k > 2$.

Examine first the case $k = 2$. For this case we shall show

Theorem 1. *There exists a sequence of random variables $\{X_i\}_{i=1}^\infty$ taking values ± 1, pairwise independent, such that S_n is bounded almost surely.*

This result should be contrasted against the fact that we know that $\mathbb{E}S_n = 0$ and $\mathbb{E}S_n^2 = n$, just like in the completely independent case. In other words, S_n for large n must have extreme "fat tail" behavior. Naturally, $M := \max_n S_n$ satisfies $\mathbb{E}M = \infty$. The example we will demonstrate is essentially the Walsh system, which is pairwise independent. We will discuss this in section 2.

Such behavior cannot occur when $k \geq 4$ for the simple reason that in this case we know that $\mathbb{E}S_n^4 = 3n^2 - 2n$ and this gives

$$\mathbb{P}(S_n^2 > M) \geq \frac{\left(\mathbb{E}S_n^2 - M\right)^2}{\mathbb{E}S_n^4} \to \frac{1}{3} \quad \forall M.$$

We could not settle $k = 3$,

Problem 1. *Is there a sequence of random variables $\{X_i\}_{i=1}^\infty$ taking values ± 1 with equal probability 3-wise independent, such that S_n is bounded almost surely.*

The higher k is, the more moments we know and we can approximate the large scale shape of S_n. However, this does not necessarily mean we can say something about $S_n \mod m$. And indeed, in section 3 we show the following

Theorem 2. *Let m and k be some natural numbers, and let $\epsilon > 0$. Then there exists a sequence of random variables $\{X_i\}_{i=1}^{\infty}$ taking values ± 1, k-wise independent, and a sequence I_j such that*

$$\mathbb{P}\left(S_{I_j} \equiv 0\ (m)\right) > 1 - \epsilon \quad \forall j. \tag{1}$$

Notice that the requirement that the condition holds only for some I_j is unavoidable, since, say, if $k \geq 10m^2$ then the distribution of $S_{I_j+10m^2}$ is approximately uniform, since $X_{I_j+1}, \dots, X_{I_j+10m^2}$ are independent.

Explicit constructions of k-wise independent 0-1 probability spaces and estimates on their sizes are the focus of several papers in combinatorics and complexity, see e.g. [3] for the first construction and [1] for a recent discussion with additional references. Sums of pairwise independent random variable were extensively studied, see e.g. [2] for some interesting examples. Thus there are already many interesting examples of pairwise independent processes. But it seems behavior modulo m was not studied.

2. PAIRWISE INDEPENDENT PROCESSES

We term the construction we use a “gray walk”, as it is based on the well-known Gray code construction in combinatorics. The nth Gray code is a Hamiltonian path on the discrete n-cube, or a listing of all 2^n binary strings of length n, starting with the string $00...0$, where every string appears exactly once, and any two consecutive strings differ in exactly one place. The construction (and hence also proof that this is possible) is done recursively, namely: To construct the nth Gray code, write down two copies of the $(n-1)$th code, where the order of the strings in the second copy is reversed, and add to the first 2^{n-1} strings a zero at the nth place, and to the second 2^{n-1} strings a one at the nth place.

The Gray codes of all orders n can be combined into an infinite Gray code, by listing them sequentially for increasing n's, and converting all strings into infinite strings by padding with zeros. The first few strings in the infinite code are:

$$\begin{array}{lll} A_0 = 000000\dots & A_1 = 100000\dots & A_2 = 110000\dots \\ A_3 = 010000\dots & A_4 = 011000\dots & A_5 = 111000\dots \\ A_6 = 101000\dots & A_7 = 001000\dots & A_8 = 001100\dots \\ & \vdots & \end{array}$$

To construct the gray walk, we now consider each string A_i as specifying a finite subset (also denoted A_i) of the natural numbers $\mathbb{N}$, where a 1 in the jth place signifies that $j \in A_i$, and define

$$X_i = \prod_{j \in A_i} \xi_j$$

where $\xi_1, \xi_2, ...$ is a seed sequence of independent ± 1 variables. It is easy to verify that the X_i's are pairwise independent. Therefore to finish theorem 1 we only need

Proposition. *The gray walk S_n is bounded almost surely. More precisely*

$$\sup_n |S_n| = 2^{\min\{j\geq 1:\xi_j=-1\}-1}+1 = \sup S_n + 2 = -\inf_n S_n$$

Proof. For simplicity we add the element $X_0 \equiv 1$, and prove that for $S'_n = \sum_{i=0}^n X_i$ we have $\sup S'_n = -\inf S'_n = 2^{\min\{j\geq 1:\xi_j=-1\}-1}$. The recursive definition of the Gray code is reflected in the path: Assume we have defined the first steps $S'_0, S'_1, S'_2, ..., S'_{2^{j-1}-1}$ of the walk, then the next 2^{j-1} steps will be the previous steps listed in reverse order, and multiplied by the random sign ξ_j. This implies that the path up to time $2^j - 1$, where j is the first value for which $\xi_j = -1$, is $0, 1, 2, 3, 4, ..., 2^{j-1}, 2^{j-1}-1, 2^{j-1}-2, ..., 3, 2, 1, 0$, and all the subsequent values are bounded between -2^j and 2^j. □

Remark. Theorem 1 holds also if one takes the strings A_i in lexicographic order, which is simpler and is also the standard ordering of the Walsh system. However, we believe the gray walk is interesting in its own right. For example, it satisfies that $\frac{X_{i+1}}{X_i} = \xi_{j(i)}$, i.e. the seed sequence is exposed directly in the quotients of the X_i's.

On the other hand, there is a vast body of knowledge about the behavior of "weighted walks" with the lexicographical order, which is simply the question when

$$\sum_{i=0}^{\infty} c_i W_i$$

converges where W_i is the Walsh system. The general philosophy is that it is similar to the behavior of the Fourier system, with the exception of the symmetry of the Fourier system which has no equivalent in the Walsh system. Of course, this is *very different* from the simple behavior of independent variables. See [5] for a survey, or the book [4].

3. Proof of theorem 2

Before embarking on the general proof, let us demonstrate the case $m = 4$ which is far simpler. Let $L > k$ satisfy $L \equiv 0 \bmod 4$. Let $\{\xi_i\}_{i=1}^{\infty}$ be a sequence of i.i.d ± 1 variables. Next define new variables by conditioning the ξ_i-s:

$$\{X_i\} := \left\{ \xi_i \,\middle|\, \prod_{b=1}^{L} \xi_{aL+b} = 1 \ \forall a = 0, 1, \ldots \right\} \quad .$$

The fact that $L > k$ clearly shows that the X_i's are k-independent. The fact that $\prod \xi_{aL+b} = 1$ shows that in each block of size L the number of -1's is even, and since $4|L$ we get that $\sum \xi_{aL+b} \equiv 0 \bmod 4$ for every block. Therefore if we define $I_j = jL$ then the condition in (1) is actually satisfied combinatorially and not only in high probability.

How can we generalize this to $m \neq 4$? The remaining ingredient in the proof is based on the following well known fact:

It is possible to construct variables with probability $\frac{1}{3}$, $\frac{1}{3}$, $\frac{1}{3}$ given a sequence of independent variables with probabilities $\frac{1}{2}$, $\frac{1}{2}$.

The algorithm is simple: take two such variables: if they give 11, "output" 1, if they give $1, -1$ "output" 2 and if they give $-1, 1$, "output" 3. If they give $-1, -1$, take a new set of two variables and repeat the process. The output is 1, 2 or 3, each with probability $\frac{1}{3}$. The time required for each step is unbounded but the probability that it is large is exponentially small.

The proof below combines these two ideas, which we nickname "generating hidden symmetries" and "simulating uniform variables" to get the result.

Proof of theorem 2, general m. We may assume without loss of generality that m is even (otherwise take $2m$ instead). Let $\lambda = \lambda(k, \epsilon)$ be some sufficiently large parameter which we shall fix later. Let $L := 2(k+1)\left\lceil \lambda m^2 \right\rceil$ where $\lceil \cdot \rceil$ stands, as usual, for the *upper* integer value.

Lemma. *For every $\mu = 0, 2, \dots, m-2$ there exists variables L^μ and Y^μ with the following properties*

(1) *L^μ is a random positive integer, L divides L^μ and $\mathbb{P}(L^\mu = aL) \le 2^{-a-1}\epsilon$ for all $a > 1$.*
(2) *$Y^\mu = \{y_i^\mu\}_{i=1}^\infty$ is a sequence of k-wise independent variables taking values ± 1 with probabilities $\frac{1}{2}$.*
(3) *$\left\{ \{y_i^\mu\}_{i=L^\mu+1}^\infty \middle| L^\mu \right\}$ are i.i.d ± 1 variables*
(4) *$L^\mu = L$ implies $\sum_{i=1}^L y_i^\mu \equiv \mu$ mod m.*

Proof. Define $N := 2\left\lceil \lambda m^2 \right\rceil$. The first step is to divide $\{\pm 1\}^N$ according to the sum modulo m, and trim the resulting sets a little so that they all have the same size. Precisely, let

$$A := \min_{j=0,2,\dots,m-2} \# \left\{ v \in \{\pm 1\}^N : \sum_{i=1}^N v_i \equiv j\ (m) \right\} \quad .$$

We note that the distribution of $\sum_{i=1}^N \pm 1$ modulo m is uniform on the set $0, 2, \dots, m-2$ with an error of $Ce^{-c\lambda}$. Therefore $2^N - \frac{m}{2}A \le C2^{N-c\lambda}$. For $j = 0, 2, \dots, m-2$, let G_j be arbitrary sets satisfying

$$G_j \subset \left\{ v \in \{\pm 1\}^N : \sum_{i=1}^N v_i \equiv \lambda\ (m) \right\} \quad |G_j| = A.$$

and let $S := \{\pm 1\}^N \setminus \bigcup G_j$.

So far we have constructed some general objects having very little to do with probability. Next, let $\{\xi_i\}_{i=1}^\infty$ be i.i.d variables taking values in ± 1, and let $\{\Xi_i\}_{i=1}^\infty$ be a division of the ξ_i's into blocks of size N: $\Xi_i := (\xi_{Ni-N+1}, \dots, \xi_{Ni})$. Let

$$B := \min \left\{ b : \# \{\Xi_i \not\in S\}_{i=1}^b = k+1 \right\} \quad .$$

We define

$$L^\mu := \left\{ L \left\lceil \frac{B}{k+1} \right\rceil \middle| \sum_{i=1}^{BN} \xi_i \equiv \mu\ (m) \right\}$$

$$Y^\mu := \left\{ (\xi_1, \xi_2, \dots) \middle| \sum_{i=1}^{BN} \xi_i \equiv \mu\ (m) \right\} \quad .$$

Properties 3 and 4 are obvious from the construction. Property 1 is a direct consequence of the estimate $\mathbb{P}(\Xi_i \in S) \le Ce^{-c\lambda}$, if only λ is large enough. Define λ so as to satisfy this condition. Therefore we need only prove property 2.

Let therefore $i_1 < \cdots < i_k$ and $\delta_1, \ldots, \delta_k \in \{\pm 1\}$ and examine the events

$$\mathcal{X} := \{\xi_{i_1} = \delta_1, \ldots, \xi_{i_k} = \delta_k\} \quad \mathcal{Y} := \left\{ \sum_{i=1}^{BN} \xi_i \equiv \mu \ (m) \right\}.$$

We know that $\mathbb{P}(\mathcal{X}) = 2^{-k}$ and we need to show that $\mathbb{P}(\mathcal{X}|\mathcal{Y}) = 2^{-k}$. Let b be some number and let $\sigma \subset \{1, \ldots, b-1\}$ be a set with $\#\sigma = b - (k+1)$ and examine the event

$$\mathcal{S} = \mathcal{S}(\sigma, b) = \{\xi_i \in S \Leftrightarrow i \in \sigma \quad \forall i \le b\} \quad .$$

The point of the proof is that the event $\mathcal{Y}$ is independent of $\mathcal{X} \cap \mathcal{S}$. This is because $\mathcal{X}$ depends only on k places, but $\mathcal{Y}$ depends on $k+1$ Ξ_i's each of which is distributed uniformly. In other words

$$\mathbb{P}(\mathcal{Y}|\mathcal{X} \cap \mathcal{S}) = \mathbb{P}(\mathcal{Y}) = \frac{2}{m}$$

or

$$\mathbb{P}(\mathcal{X} \cap \mathcal{S}|\mathcal{Y}) = \frac{\mathbb{P}(\mathcal{X} \cap \mathcal{S} \cap \mathcal{Y})}{\mathbb{P}(\mathcal{Y})} = \mathbb{P}(\mathcal{X} \cap \mathcal{S})$$

which gives

$$\mathbb{P}(\mathcal{X}|\mathcal{Y}) = \sum_{\sigma,b} \mathbb{P}(\mathcal{X} \cap S|\mathcal{Y}) = \sum_{\sigma,b} \mathbb{P}(\mathcal{X} \cap \mathcal{S}) = \mathbb{P}(\mathcal{X}) = 2^{-k} \quad . \qquad \square$$

Returning to the proof of the theorem, we let

$$\{Y^{\mu,\nu}, L^{\mu,\nu}\}_{(\mu,\nu) \in \{0,2,\ldots,m-2\} \times \mathbb{N}}$$

be an independent family of couples of variables constructed using the lemma. We now construct our sequence X_i inductively as follows: assume that at the νth step we have defined $X_1, \ldots, X_\rho$ where $\rho = \rho(\nu)$ is random. Define $\mu = \mu(\nu) \equiv -\sum_{i=1}^{\rho} X_i \bmod m$ and then define $X_{\rho+1}, \ldots, X_{\rho+L^{\mu,\nu}}$ using

$$X_{\rho+i} = y_i^{\mu,\nu}$$

and $\rho(\nu+1) = \rho(\nu) + L^{\mu,\nu}$. This creates a sequence of ±1 variables. To see (1), define $r(a) = \max\{\nu : \rho(\nu) < aL\}$ and properties 1 and 4 of the lemma will give that

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{aL} \epsilon_i \not\equiv 0 \ (m)\right) &\le \sum_{b=1}^{\infty} \mathbb{P}\big(\rho(r(a)) = (a-b)L \text{ and } L^{\mu(r(a)),r(a)} \ge L \max\{b,2\}\big) \\ &\le \sum_{b=1}^{\infty} \sum_{c=\max\{b,2\}}^{\infty} \mathbb{P}(L^{\mu(r(a)),r(a)} = cL) \le \sum_{b,c} 2^{-c-1}\epsilon = \epsilon \quad .\end{aligned}$$

Therefore we need only show that the X_i's are k-wise independent. While being strongly related to the k-wise independence of the $y_i^{\mu,\nu}$ it is not an immediate consequence of it because of the dependence between the $y_i^{\mu,\nu}$ and the $L^{\mu,\nu}$'s.

We prove that the X_i's are k-wise independent inductively. Let l and I be some integers and assume that we have shown that $X_{i_1}, \ldots, X_{i_m}$ are independent for every $m < l$ and for $m = l$ and $i_1 < I$. Let $I = i_1 < i_2 < \cdots < i_l$ and let $\delta_1, \dots, \delta_l \in \{\pm 1\}$. Examine $L^{0,1}$ and define

$$n = n(L^{0,1}) = \#\{m : i_m \le L^{0,1}\} \quad .$$

The induction hypothesis, together with the independence of the various $Y^{\mu,\nu}$'s shows that $X_{i_{n+1}}, \dots, X_{i_l}$ are i.i.d ± 1 variables: if $n > 0$ then this follows from the induction hypothesis for $m' := m - n$ while if $n = 0$ then it follows from the induction hypothesis for $m' = m$ and $I' := I - L^{0,1}$. Property 3 of the lemma shows that $y^{0,1}_{i_{n+1}}, \dots, y^{0,1}_{i_l}$ are i.i.d ± 1 variables. Below n we have simple equality:

$$X_{i_m} = y^{0,1}_{i_m} \quad \forall m \le n.$$

Therefore

$$\mathbb{P}(X_{i_1} = \delta_1, \dots, X_{i_l} = \delta_l | L^{0,1}) = \mathbb{P}(y_{i_1} = \delta_1, \dots, y_{i_l} = \delta_l | L^{0,1})$$

and summing over $L^{0,1}$ we get the required result:

$$\mathbb{P}(X_{i_1} = \delta_1, \dots, X_{i_l} = \delta_l) = \mathbb{P}(y_{i_1} = \delta_1, \dots, y_{i_l} = \delta_l) = 2^{-l}$$

which finishes the proof of the theorem. □

Remarks.

(1) It is easy to generalize the theorem to force $\sum_{i=1}^{I_j} \epsilon_i$ to have almost arbitrary distributions. The only obstacle is the even-odd symmetry of the walk. Thus if m is odd, any distribution whatsoever on the congruence classes modulus m might be achieved, while if m is even, any distribution supported on the set of even or odd congruence classes can be achieved.

(2) It is easy to see from the proof that $L \approx km^2 \log(k/\epsilon)$. In other words, the precision is asymptotically exponential in the stepping of the subsequence I_j. We remind the reader a fact that was discussed in the introduction: if $k > m^2 \log 1/\epsilon$ then necessarily $I_{j+1} - I_j > cm^2 \log 1/\epsilon$ because immediately after I_j we get a sequence of independent variables.

(3) Another direction in which k-wise independent random variables may be very different from truly independent variables is their entropy. It is known that the entropy of n such variables may be as low as $k \log n$, see [3] and [1]. Our pairwise example is optimal in this respect: n truly random bits create 2^n pairwise independent random variables. The example above is far from it: the entropy is linear in the number of bits. However, it turns out that it is possible to create an example of k-wise independent variables with low entropy satisfying the conditions of theorem 2. The example, while amusing, if a bit off topic hence we will skip it.

Theorem 3. *Let m and k be some natural numbers. Then there exists a sequence of random variables $\{\epsilon_i\}_{i=1}^{\infty}$ taking values ± 1, k-wise independent, and a sequence I_j such that*

$$\sum_{i=1}^{I_j} \epsilon_i \equiv 0 \; (m) \textit{ for } j \textit{ sufficiently large with probability } 1.$$

The proof is similar to that of theorem 2, but using a different L in each step. However there are some additional technical subtleties. Essentially, if in the main lemma of the previous theorem we defined a variable Y^{μ} where the parameter μ was used to "return to sync" the series if ever the congruence is no longer 0, here we would need variables $Y^{\mu,\tau}$ where τ is some additional parameter needed to "return to sync" in the L domain. While being only moderately more complicated, we felt it better to present the simpler proof of the previous theorem.

4. Further related problems

We end by suggesting further problems related to k-wise independent random variables.

4.1. **The random sign Pascal triangle.** Let $\xi^{+}_{n,k}, \xi^{-}_{n,k}$ be a family of i.i.d random (1/2-1/2) signs for $-1 \le k \le n+1$. Define random variables $X_{n,k}$ for $-1 \le k \le n+1$ by the recurrence

$$X_{0,0} = 1, \quad X_{n,-1} = 0, \quad X_{n,n+1} = 0,$$

$$X_{n,k} = \xi^{+}_{n-1,k-1} X_{n-1,k-1} + \xi^{-}_{n-1,k} X_{n-1,k} \quad (0 \le k \le n)$$

Problems: Study the behavior of the central coefficient $X_{2n,n}$. Find an interpretation of the triangle as "percolation with interference" where two closed gates cancel each other out ("quantum percolation"?) Study the behavior of the non-central coefficients.

4.2. **2D Percolation.** What about other functions of k-wise independent random variables? For example, instead of considering Bernoulli percolation, the random variables determining the configuration are only k-wise independent. Here is an "unimpressive" example: k-wise independent bond percolation on an $n \times n(k+1)$- box in $\mathbb{Z}^2$ such that the probability to have a crossing of the narrow side is 1. Of course, even in usual percolation the probability[1] is $1 - e^{-ck}$ so that the improvement is not exciting.

The construction is as follows: divide into $k+1$ disjoint boxes of size $n \times n$. Take the usual independent percolation, conditioned such that the number of boxes with crossings is odd. This is like conditioning $k+1$ Bernoulli variables to have an odd sum, so that you get a k-wise independent structure. If the number is odd, then it is non zero. Can it be improved, say to guarantee crossing of the $n \times n$-box?

References

[1] N. Alon, and J. Spencer, *The probabilistic method.* Second edition. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[2] S. Janson, *Some pairwise independent sequences for which the central limit theorem fails*, Stochastics 23/4 (1988), 439–448.

[3] A. Joffe, *On a set of almost deterministic k-independent random variables*, Ann. Probability 2/1 (1974), 161–162.

[4] F. Schipp, William R. Wade and P. Simon (with assistance by J. Pàl), *Walsh Series: An Introduction to Dyadic Harmonic Analysis*, Adam Hilger Ltd., Bristol and New York, 1990.

[1] This is obvious if you are willing to use the conformal invariance of percolation. Otherwise, this fact can be derived from Russo-Seymour-Welsh theory.

[5] William R. Wade, *Dyadic harmonic analysis*, In Contemporary Mathematics 208, Harmonic analysis and nonlinear differential equations, Riverside CA 1995, 313–350.